

マルコフ連鎖モデルによる文節かな漢字変換候補の絞り込み 方法

荒木 哲郎⁺ 池原 悟⁺⁺ 芳永 寛司⁺⁺⁺ 真田 陽一⁺

⁺福井大学

⁺⁺NTT 情報通信網研究所

⁺⁺⁺松下電器産業

Abstract

本論文では、文節境界が与えられた場合、文節切りされたべた書き音節表記文節から、かな漢字変換によって生じる漢字かな交じり文節候補を絞り込む問題に対して、文節内及び文節間の漢字かな文字及び品詞情報のマルコフ連鎖モデルを用いる方法を提案する。実際に、43万語単語辞書を参照して、新聞記事77日分の統計データを用いた実験を行った結果、次のような知見を得た。

1. かな漢字変換において、文節を構成する単語候補の組み合わせは、文節内の単語分割数が最小となる数より1つ大きい場合までを含めることにより、約95%の正解率が得られることがわかった。
2. 文節内の漢字かなコストにより漢字かな候補を順序づけ、さらに文節内の品詞コストを足切り値によって評価する方法により、特に標本外データの場合、10位内正解候補内約4割を、削減できることがわかった。
3. 文節内と文節間の漢字かなコストとの組み合わせ法により、最終的に第一位正解率が、83.7%~98.2%になることがわかった。

A Method of Deciding "Bunsetsu" Candidates of "kanji-kana" Strings Using Markov Models

Tetsuo ARAKI⁺ Satoru IKEHARA⁺⁺ Hiroshi Yoshinaga⁺⁺⁺ Youichi Sanada⁺

⁺Fukui University

⁺⁺NTT Network Information Systems Laboratories

⁺⁺⁺Matshita Electric Industrial Company Limited

There are many reseaches on the method which translates the non-segmented "Kana" sentences into the "kanji-kana" sentences. However, the amount of computer memories required for the translating processing explodes in many times, because the number of the combinations of candidates for "kanji-kana" words grows rapidly in propotion to te increasing of the length of the sentence. The memory explosion can be prevented if a sentences is separated into "bunsestu". Up to now, an useful mehod for finding and correcting the provisional boundaries of "bunsestu" using 2nd-order Markov model has been proposed. This paper proposes a method of deciding the most provable candidate of "kanji-kana" strings translated from the non-segmented "kana bunsetsu" described above, using Markov models of "kanji-kana" characters and word categories.

I. はじめに

日本語を計算機に入力する方法としては、現在区切りを一切入れないべた書き文入力方式が主流となっている。べた書きかな文の場合に、総当たり法でかな漢字変換等の処理により生成される、あらゆる単語候補列の組み合わせを考慮して解析を行うと、一般に探索木が爆発する問題が生じる。従来、かな文については最長一致法による方法 [1]、文節数最小法 [2]、前後の接続文字を利用した方法 [3]、格文法を用いる方法 [4]、連語解析を用いる方法 [5]、単語共起の関係を用いる方法 [6] 等があるが、同音語による曖昧さと分かち書き処理の曖昧さを同時に解決しなければならず、現在のところではまだ十分な精度を得るには至っていない。

本論文では、[11]-[14] において提案されている仮文節境界の推定並びに補正方法によって求められる文節境界を前提とし、正しく与えられたべた書きかな文節より、かな漢字変換によって得られる膨大な漢字かな候補列を絞り込む問題に対して、マルコフ連鎖モデルを用いた音節認識候補列の絞り込み法 [9] 及び漢字かな候補絞り込み法 [10] を拡張し、文節内及び文節間の漢字かな文字及び品詞情報のマルコフ連鎖モデルを用いる方法を提案する。

実際に、43万語単語辞書を参照して、新聞記事77日分の統計データを用いた実験行ってその有効性を評価する。更に、マルコフ連鎖確率による絞り込みの方法として、漢字かな文字列で正規化する場合としない場合の評価ならびに音節漢字変換処理の際に用いる文節内の単語分割数と正解率の関係についても実験結果を示す。

II. 諸定義

【定義1】日本語文 $S(X) = x_1 x_2 \dots x_l$ において、すべての文字 x_i が音節文字であるときを音節文、または全ての x_i が {漢字、ひらかな、カタカナ、英数字、記号等} のいずれかの文字で表されるとき $S(X)$ を漢字かな交じり文と呼び、 l を文 $S(X)$ の長さと呼ぶ。 $S(X)$ の中の連続した文字の部分列 $x_i x_{i+1} \dots x_{i+r}$ が自立語 (詞) = (名詞、動詞、形容詞、形容動詞、副詞、連体詞、接続詞、感動詞、形式名詞) と付属語 (辞) = (助詞、助動詞、接辞) から構成される一つの単位を文節と呼び、これを単に $B(X) = x_1 x_2 \dots x_n$ と表す。但し、 $B(X)$ には自立語は必ず含まれるが、付属語は必ずしも含まれるとは限らない。

音節文節に対して、それぞれ以下のように単語辞書引きして、候補が存在した単語を順次接続して単語境界の整合がとれた文節候補列の集合を、音節候補列から生成された漢字かな文節候補の集合と呼び Γ で表す。ここで音節表記文字は、日本語音声出力 [8] で用いた 110 音節を用いる。

【定義2】単語辞書は音節列をキー見出しとして、漢字かな表記の単語を読み出せるものとする。文節候補 $B = x_1 x_2 \dots x_n$ の中の部分音

節列 $c = x_i x_{i+1} \dots x_{i+r}$ をキーとして辞書引きを行い、その候補が存在したとき、音節列 c をキーとする漢字かな単語候補 $\gamma(c)$ の集合を $\Gamma(c)$ と表す。 B を互いに重なり部分がない音節部分に分割 $B = c_1 + c_2 + \dots + c_w$ したとき、全ての c_j ($1 \leq j \leq w$) に対して単語辞書に各単語候補 $\gamma(c_j)$ が全て存在する場合、これらの単語候補を連結した候補列 $\gamma(c_1) \gamma(c_2) \dots \gamma(c_w)$ を漢字かな交じり文節候補と呼ぶ。

その時の B の音節部分列への分割数 w を単語分割数と呼び、文節における最小の単語分割数を最小分割数と呼ぶ。

【定義3】標本内の各文節 B の漢字かな交じり表記を、 $B(X) = x_1 x_2 \dots x_n$ 、また B の品詞表記を、 $B(Y) = y_1 y_2 \dots y_s$ とする。このとき漢字かな文字 x_i 及び品詞 y_j に対して定義される文節内の m 重マルコフ連鎖確率 $p_i = (x_i | x_{i-m} \dots x_{i-1})$ 及び $q_j = (y_j | y_{j-m} \dots y_{j-1})$ の集合を、それぞれ $P_{(in)}^{(m)}(X)$ 及び $P_{(in)}^{(m)}(Y)$ と表す。但し、 x_k で $k \leq 0$ または $k \geq n$ 及び y_k で $k \leq 0$ または $k \geq s$ ならば、 x_k 及び y_k を文節境界を表す空白文字 b とする。このとき $P_{(in)}^{(m)}(X)$ 及び $P_{(in)}^{(m)}(Y)$ を用いて、次のように定義される文節 B の確からしさ $C_{(in)}(X)$ 及び $C_{(in)}(Y)$ をそれぞれ、文節内漢字かなコスト及び文節内品詞コストと呼ぶ。

$$(1) C_{(in)}(X) = - \sum_{i=1}^{n+m} \log_2 p_i$$

$$(2) C_{(in)}(Y) = - \sum_{j=1}^{s+m} \log_2 q_j$$

本論文では、文節内の m 重マルコフ連鎖確率の次数 m は、2重と3重を取り扱うことにする。

【定義4】標本内の各文節及びその直前、直後の文節漢字かな表記をそれぞれ次のように $B^{(i)}(X)$, $B^{(i-1)}(X)$, $B^{(i+1)}(X)$ と表す。

$$B^{(i)}(X) = x_1^{(i)} x_2^{(i)} \dots x_m^{(i)}$$

$$B^{(i-1)}(X) = x_1^{(i-1)} x_2^{(i-1)} \dots x_l^{(i-1)}$$

$$B^{(i+1)}(X) = x_1^{(i+1)} x_2^{(i+1)} \dots x_n^{(i+1)}$$

また各文節の直前、直後の文節の品詞表記を、それぞれ次のように $B^{(i)}(Y)$, $B^{(i-1)}(Y)$, $B^{(i+1)}(Y)$ と表す。

$$B^{(i)}(Y) = y_1^{(i)} y_2^{(i)} \dots y_s^{(i)}$$

$$B^{(i-1)}(Y) = y_1^{(i-1)} y_2^{(i-1)} \dots y_s^{(i-1)}$$

$$B^{(i+1)}(Y) = y_1^{(i+1)} y_2^{(i+1)} \dots y_s^{(i+1)}$$

このとき漢字かな表記文節 $B^{(i)}(X)$ 及び品詞表記文節 $B^{(i)}(Y)$ の直前、直後の文節に対して、次のように定義される $P_{(fw-bet),i}^{(m)}$, $P_{(bw-bet),i}^{(m)}$ および

$q_{(fw-bet),i}^{(m)}$, $q_{(bw-bet),i}^{(m)}$ をそれぞれ前、後文節間の m 重漢字かなマルコフ連鎖確率及び前、後文節間の m 重品詞マルコフ連鎖確率と呼び、その集合を $P_{(fw-bet)}^{(m)}(X)$, $P_{(bw-bet)}^{(m)}(X)$

また、 $P_{(fw-bet)}^{(m)}(Y)$ 及び $P_{(bw-bet)}^{(m)}(Y)$ と表

す。

(1) $P_{(fw-bet),i} = (x_i | \hat{x}_{i-m} \cdots \hat{x}_{i-1})$, (但し、 $1 \leq i \leq m$ で x_i は $B^{(i)}(X)$ の文字である。また \hat{x}_j で $j \leq 0$ ならば \hat{x}_j は、 $B^{(i-1)}(X)$ の文字であり、 $j \geq 1$ ならば \hat{x}_j は、 $B^{(i)}(X)$ の文字である)

(2) $P_{(bw-bet),i} = (x_i | \hat{x}_{i-m} \cdots \hat{x}_{i-1})$, (但し、 $1 \leq i \leq m$ で x_i は $B^{(i+1)}(X)$ の文字である。また \hat{x}_j で $j \leq 0$ ならば \hat{x}_j は、 $B^{(i)}(X)$ の文字であり、 $j \geq 1$ ならば \hat{x}_j は、 $B^{(i+1)}(X)$ の文字である)

(3) $q_{(fw-bet),i} = (y_i | \hat{y}_{i-m} \cdots \hat{y}_{i-1})$, (但し、 $1 \leq i \leq m$ で y_i は $B^{(i)}(Y)$ の文字である。また \hat{y}_j で $j \leq 0$ ならば \hat{y}_j は、 $B^{(i-1)}(Y)$ の文字であり、 $j \geq 1$ ならば \hat{y}_j は、 $B^{(i)}(Y)$ の文字である)

(4) $q_{(bw-bet),i} = (y_i | \hat{y}_{i-m} \cdots \hat{y}_{i-1})$, (但し、 $1 \leq i \leq m$ で y_i は $B^{(i+1)}(Y)$ の文字である。また \hat{y}_j で $j \leq 0$ ならば \hat{y}_j は、 $B^{(i)}(Y)$ の文字であり、 $j \geq 1$ ならば \hat{y}_j は、 $B^{(i+1)}(Y)$ の文字である)

このとき $P_{(fw-bet)}^{(m)}(X)$ 、 $P_{(bw-bet)}^{(m)}(X)$ 及び、 $P_{(fw-bet)}^{(m)}(Y)$ 、 $P_{(bw-bet)}^{(m)}(Y)$ を用いて、次のように定義される文節 B の確からしさ $C_{(fw-bet)}(X)$ 、 $C_{(bw-bet)}(X)$ 及び、 $C_{(fw-bet)}(Y)$ 、 $C_{(bw-bet)}(Y)$ をそれぞれ、前文節間の漢字かなコスト、後文節間の漢字かなコスト及び、前文節間の品詞コスト、後文節間の品詞コストと呼ぶ。

- (i) $C_{(fw-bet)}(X) = - \sum_{i=1}^{n+m} \log_2 P_{(fw-bet),i}$
- (ii) $C_{(bw-bet)}(X) = - \sum_{i=1}^{n+m} \log_2 P_{(bw-bet),i}$
- (iii) $C_{(fw-bet)}(Y) = - \sum_{i=1}^{n+m} \log_2 q_{(fw-bet),i}$
- (iv) $C_{(bw-bet)}(Y) = - \sum_{i=1}^{n+m} \log_2 q_{(bw-bet),i}$

本論文では、前、後文節間の m 重マルコフ連鎖確率次数は、 $m = 2$ とする。音節漢字変換によって得られる漢字かな文節候補列の集合 Γ に対して、定義 3 で与えられた文節内の漢字かなコスト及び品詞コストまた、定義 4 の文節間の漢字かなコスト及び品詞コストが最小な候補を、最尤な候補として絞り込む。

III. マルコフ連鎖モデルによる漢字かな文節候補の絞り込み方法

ここでは、音節漢字変換によって得られる漢字かな文節候補の集合 Γ の中から、文節内マルコフ連鎖及び、文節間マルコフ連鎖情報を用いて絞り込む方法を述べる。

【方法 1】漢字かな文節候補の集合 Γ を、文節内漢字かなコスト値 $C_{(in)}(X)$ によって絞り込む方法を、IK法と呼ぶ。¹ また IK法に、文

¹ 本論文では、IK法におけるマルコフ連鎖確率の次数 m は、2重と3重を考える。

節内の品詞コスト値 $C_{(in)}(Y)$ が足切り値 T より、小さい条件を付加することによって絞り込む方法を、INK法と呼ぶ。²

【方法 2】漢字かな文節候補の集合 Γ を、IK法で絞り込む際に、特に文節内漢字かなコスト値 $C_{(in)}(X)$ を、各漢字かな文節の文字列長で割った値 (正規化と呼ぶ) によって絞り込む方法を、INK法と呼ぶ。

【方法 3】漢字かな文節候補の集合 Γ を、文節内漢字かなコスト値 $C_{(in)}(X)$ 及び、前、後文節間の漢字かなコストを組み合わせた方法によって絞り込む場合、

(1) 両者のコストを加算した和が最小となるもので評価するタイプを加算タイプと呼び、ADD-IK-FK法 (前文節間のみ)、ADD-IK-BK法 (後文節間のみ)、ADD-IK-FBK法 (前後文節間) と呼ぶ。

(2) 文節内漢字かなコスト値で順序づけられた候補の正しさを、前、後文節間の漢字かなコスト値が、足切り値 T より大きいかな否かで評価するタイプを足切りタイプとよび、CHK-IK-FK法 (前文節間のみ)、CHK-IK-BK法 (後文節間のみ)、CHK-IK-FBK法 (前後文節間) と呼ぶ。

【方法 4】漢字かな文節候補の集合 Γ を、IK法並びに、前、後文節間の品詞コスト値を組み合わせた方法によって絞り込む場合、

(1) 両者のコストを加算した和が最小となるもので評価するタイプを加算タイプと呼び、ADD-IK-FH法 (前文節間のみ)、ADD-IK-BH法 (後文節間のみ)、ADD-IK-FBH法 (前後文節間) と呼ぶ。

(2) 文節内漢字かなコスト値で順序づけられた候補の正しさを、前、後文節間の漢字かなコスト値が、足切り値 T より大きいかな否かで評価するタイプを足切りタイプとよび、CHK-IK-FH法 (前文節間のみ)、CHK-IK-BH法 (後文節間のみ)、CHK-IK-FBH法 (前後文節間) と呼ぶ。

IV. 実験条件

1. 入力データ

- (a) 文の種類：新聞記事
- (b) 標本データとの関係：標本内データ (マルコフ連鎖確率を求めた統計データと同じ文) と標本外データ
- (c) 字種：べた書き音節文節

² 一般には、漢字かなコストと品詞コストの組み合わせ方は、両者を加算するタイプも考えられるが、品詞マルコフ連鎖確率のエントロピーは、漢字かな文字のエントロピーよりも大きいために、単純に加算することにより漢字かなコスト単独の場合よりも悪くなる事に注意する。

- (d) 総文章数 (標本内データ): 218 - 1000 文節
 総文章数 (標本外データ): 263 - 1000 文節)

2. 使用辞書

- (a) 漢字かな候補の生成: 43万語の単語辞書
 (b) 漢字かな候補の絞り込み: 漢字かなマルコフ連鎖確率情報 (文節内の2重及び3重マルコフ連鎖確率と文節間の2重マルコフ連鎖確率)、品詞マルコフ連鎖確率情報、漢字かなの2重マルコフ連鎖確率 (新聞記事77日分の統計データにより作成)

V. 実験結果

べた書き音節文節より、かな漢字変換によって得られる漢字かな交じり文節候補を、III. で述べた方法1~4によって絞り込みを行った結果、全体として第一位正解率が83.7% (標本外データ) ~ 98.2% (標本内データ) の正解率が得られることがわかった。以下にそれぞれの場合の考察を示す

[1] 音節漢字変換における一文節内の単語分割数 (最小分割数)

音節文節から43万語の単語辞書引きをして得られる漢字かな文節候補には、文節内の単語分割数が種々なものが含まれている。その中で最も小さい単語分割数 (最小分割数) を基準に分割数と文節候補の正解率の関係を、図1に示す。同図より、最小分割数+1の所で、ほぼ飽和していることがわかる。

[2]

IK法における2重マルコフと3重マルコフの効果

文節内の漢字かなコスト用いて評価するIK法における、2重マルコフと3重マルコフの効果、図2に示す。同図より、漢字かな文字の3重マルコフ連鎖確率を用いる効果は小さく、2重マルコフ連鎖確率を用いた絞り込みで十分であることがわかる。

[3] 文節内の漢字かなコストの正規化の影響

漢字かな文節候補の集合 Γ を、IK法で絞り込む場合と、文節内漢字かなコスト値 $C_{(in)}(X)$ を、各漢字かな文節の文字列長で正規化するINK法で絞り込む場合を比較した結果を、図3に示す。

同図より、正規化するINK法より、正規化をしないIK法の方が、若干良くなる結果となる。これは日本語の文節としての漢字かな文字列を決定

する場合には、文字列長の情報も重要な要素であることを示している。

[4] 文節内の漢字かなコストと品詞コストの組み合わせ効果

漢字かな文節候補の集合 Γ を、文節内漢字かなコスト値 $C_{(in)}(X)$ によって絞り込むIK法に、文節内の品詞コスト値 $C_{(in)}(Y)$ が足切り値 T より、小さい条件を付加することによって絞り込むIKH法の効果を図4に示す。

同図より、品詞コスト値を組み合わせた効果は、IK法によって得られた10位内正解候補のうち、誤りの候補を平均約4割削減できる効果がある。

[5] 文節内と文節間の漢字かなコストを組み合わせた効果

漢字かな文節候補の集合 Γ を、文節内漢字かなコスト値 $C_{(in)}(X)$ 及び、前、後文節間の漢字かなコストを組み合わせたADD-IK-FBK法及びCHK-IK-FBK法 (前後文節間) による絞り込みの結果を図5に示す。

同図より、加算タイプの組み合わせ方法が、足切り方法よりも有効であることがわかり、第一位正解率がIK法より約2~3%向上し、83.7% (標本外データ) ~ 98.2% (標本内データ) に達することがわかった。また前文節間と後文節間とのコスト差異はほとんどみられないことがわかった。

[6] 文節内の漢字かなコストと文節間の品詞コストを組み合わせた効果

漢字かな文節候補の集合 Γ を、文節内漢字かなコストと文節間品詞コストを組合わせた、ADD-IK-FBH法、CHK-IK-FBH法 (前後文節間) によって絞り込む効果を図6に示す。

同図より、IK法に比べて加算タイプはほとんど効果はなく、足切りタイプで1割程度の候補削減効果がみられた。

IK法及びADD-IK-FBK法による漢字かな文節候補の絞り込み結果の例を、図7に示す。

VI. おわりに

本論文では、マルコフ連鎖モデルを用いて漢字かな交じり文節候補の正解率を向上させる方法を検討する研究の一環として、(1) 漢字かな文字及び品詞情報、(2) 文節内と文節間情報、(3) 2重マルコフと3重マルコフ連鎖情報による、絞り込み効果を調べそれらの持つ情報の有効性を定量的に把握した。実際に、43万語単語辞書を参照して、新聞記事77日分の統計データを用いた実

験を行った結果、次のような知見を得た。

1. かな漢字変換において、文節を構成する単語候補の組み合わせは、文節内の単語分割数が最小となる数より1つ大きい場合までを考慮することにより、95%の正解率が得られれば飽和していることがわかった。
2. 漢字かな文字のマルコフ連鎖確率を用いたかな候補絞り込みにおいては、各文節候補に対して付与されるマルコフ連鎖確率の値(コスト)を、文節長で正規化するよりもしない方が、若干正解率が良いことがわかった。
3. 漢字かな文字の3重マルコフ連鎖確率を用いる効果は小さく、2重マルコフ連鎖確率を用いた絞り込みで十分であることがわかった。
4. 文節内の漢字かな文字のコストと文節内の品詞コストの組み合わせ方は、漢字かなコストにより漢字かな候補を順序づけ、品詞コストを足切り値によって正しさを検証する方法が有効であり、特に標本外データの場合、漢字かなコストで絞り込まれた10位内候補内約4割を、削減する効果があった。
5. 文節内と文節間の漢字かなコストとの組み合わせ法により、最終的に第一位正解率が、83.7%~98.2%になることがわかった。

今後の課題としては、構文解析情報及び意味解析情報を用いた文候補絞り込みの研究が上げられる。

参考文献

1. 牧野, 木沢: "べた書き文の分かち書きとかな漢字変換—二文節最長一致法による分かち書き", 情処論, 20, 4, pp. 337-245 (1979)
2. 吉村, 日高, 吉田: "文節数最小法を用いたべた書き日本語文の形態素解析", 情処論, 24, pp. 40-46 (1983)
3. 栃内, 伊藤, 鈴木: "前後接続文字を利用した同音語選択機能を有するかな漢字変換システム", 情処論, 27, 3, pp. 313-320 (1986)
4. 大島, 阿部, 湯浦, 武市: "格文法による仮名漢字変換の多義解消", 情処論, 27, 7, pp. 679-687 (1986)
5. 本間, 山階, 小橋: "連語解析を用いたべた書きかな漢字変換", 情処論, 27, 11, pp. 1062-1067 (1986)
6. 内山, 板橋: "共起関係を利用した日本語複合名詞の分割", 情処NL研究会, 91-7, pp. 57-64 (1992)

7. 宮崎: "係り受け解析を用いた複合語の自動分割", 情報処理, Vol. 25, 6, pp. 970-979, 623-656 (1984)
8. 宮崎, 大山: "日本語音声出力のための言語処理方式", 情報処理, Vol. 27, 11, pp. 1053-1061 (1986)
9. 荒木, 村上, 池原: "2重音節マルコフモデルによる日本語の文節音節認識候補の曖昧さの解消効果", 情報処理, Vol. 30, 4, pp. 467-477 (1989)
10. 村上, 荒木, 池原: "日本語文音節入力に対して2重マルコフ連鎖モデルを用いた漢字かな交じり候補の抽出精度", 信学論, Vol. J75-DII, pp. 11-20 (1992)
11. 土橋, 荒木, 池原: "2重マルコフ連鎖確率を用いたべた書き日本語文の文節境界推定", 信学会春期大会, Vol. 6, No. D-102, pp. 104 (1993)
12. 荒木, 池原, 土橋: "2重マルコフ連鎖モデルを用いたべた書き日本語文の文節先頭位置推定法の評価", 情処NL研究会, Vol. 94-8, pp. 55-61 (1993)
13. 荒木, 池原, 土橋: "べた書きかな文の仮文節境界の補正方法", 情処自然言語処理研究会, 98-1, pp. 1-7 (1993)
14. T. Araki, S. Ikehara and J. Tutihasi: "A New Method of Finding Provisional Boundaries of "bunsestu" using 2nd-Order Markov Model", 2nd IEEE Int. Workshop on Robot and Human Communication, pp. 114-119 (1993)

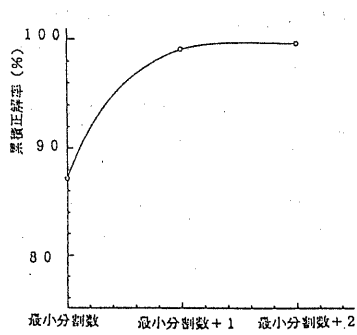


図1

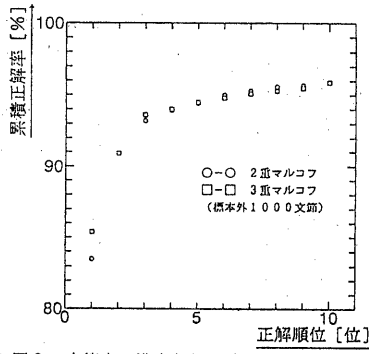


図2

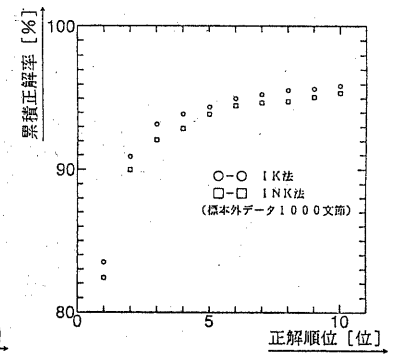
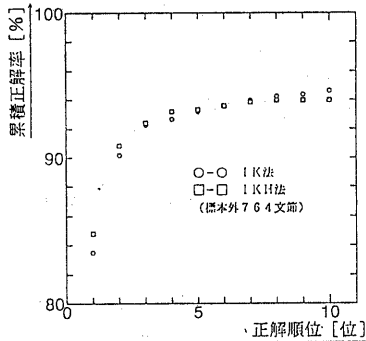
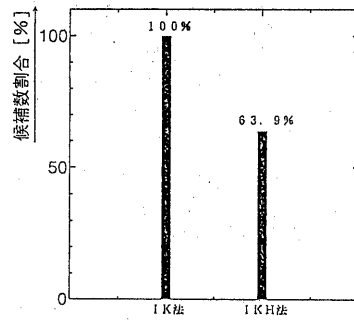


図3

図1 文節内の単語分割数と累積正解率の関係 図2 文節内の漢字かな3重マルコフ連鎖確率による効果 図3 文節内の漢字かなコストの正規化の有無による効果

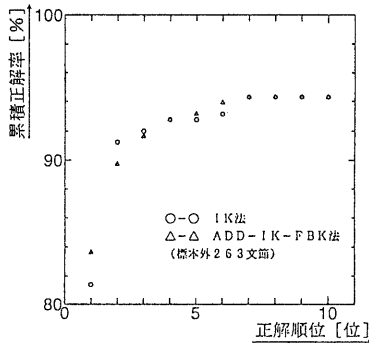


(a) 累積正解率

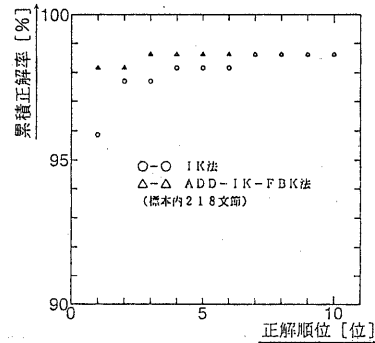


(b) 品詞コストによる足切りを行った場合の漢字変換候補の削減効果

図4 文節内漢字かなコストと品詞コストの組み合わせによる効果



(a) 標本外



(b) 標本内

図5-1 文節内と文節間の漢字かなコストを組み合わせた効果(加算タイプ)

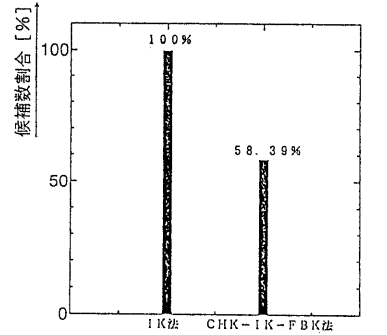
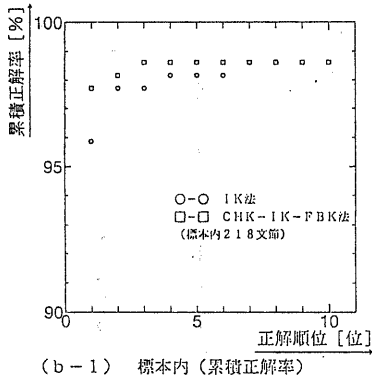
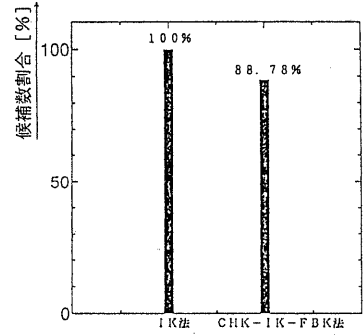
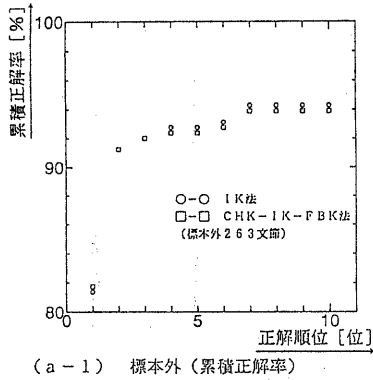


図 5-2 文節内と文節間の漢字かなコストを組み合わせさせた効果 (足切りタイプ)

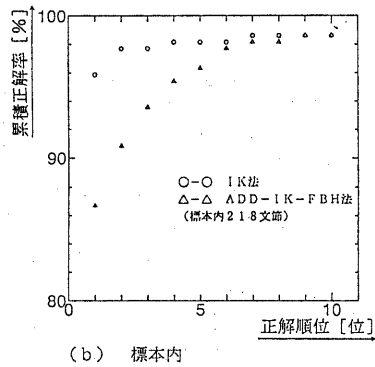
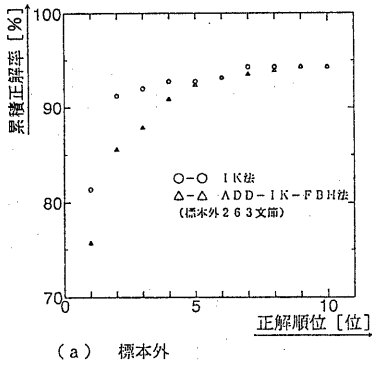
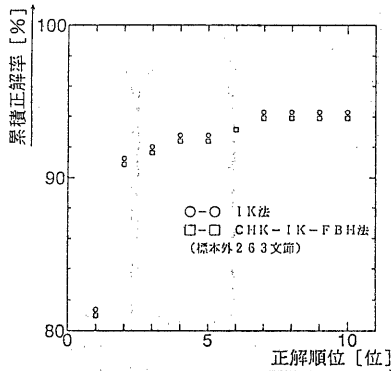
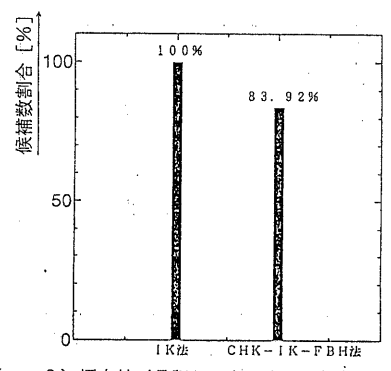


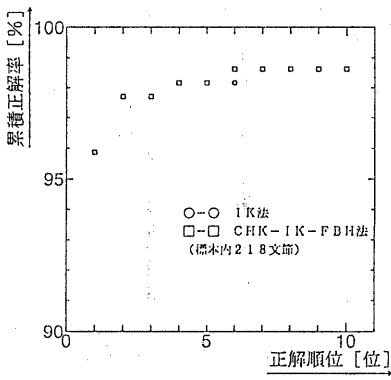
図 6-1 文節内の漢字かなコストと文節間の品詞コストを組み合わせさせた効果 (加算タイプ)



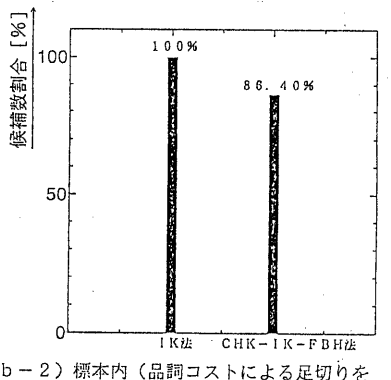
(a-1) 標本外 (累積正解率)



(a-2) 標本外 (品詞コストによる足切りを行った場合のかな漢字変換候補の削減効果)



(b-1) 標本内 (累積正解率)



(b-2) 標本内 (品詞コストによる足切りを行った場合のかな漢字変換候補の削減効果)

図6-2 文節内の漢字かなコストと文節間の品詞コストを組み合わせた効果 (足切りタイプ)

関する和らげるのに役立つと国経済のこの規定を在日外銀に高まると自己資本はその進出外銀の本店店合計の自己資本を計算基準に法体系を海外でワラ似た接近するまでワラ上がればすぐ株式に転換される増加がなだらかという特徴がワラントと海外で発行すれば外貨建て債権の多い輸出企業に転換で社債が減税問題に関する首相の発言はある都道府県にわざわざ可能な技術を以前から最も脚光を浴びている現場で砂漠の防空システムの指示を(4→3位)是正に取り組み韓国の産業経済をする期待と改造からする首相の認識の現実的て外相の具体的圧迫しないように無意味ではとるべきテコ入れに担当する政務秘書官の急逝で次も負けるときはいろいろな形のカプセルを土地の形や周囲のなる今後建設費の高騰にある小型店舗を増やす(2→1位)傾向が環境が少なくなったりした変化に備えた店舗戦略が支店長から景気の大型の銀行窓口販売は昨年証券界の意見がもめた証券取引法上の買(2→1位)動きが年金支給に何ら影響はひとつ見逃すことのできない決定した直後のいずれが適当かの判断が見直しも懇談会で検討が現状に抱いているのも限られた予算の中で今後の教育の重点を示す時期に設置する外銀数は百を借り手の外銀を利用することが波をダウンに原因は各国で足をとられた前途は政権は苦戦を覚悟せねば乗せる青写真作りをひとつは財政赤字を経済の支出をどう(2→1位)この問題で大統領とその不穏当な発言を推測と戦後わたって質問されるのが嫌いなので動員して議員の泣きどころを逆転勝利を大統領は死(7→1位)だり失うことは彼の地位を脅かす再びしたので転手技術を応用して開発中の鉄浴ガス化を周辺機器工場を品目は産油国はこぞって石油製品のトン当たり取り巻く環境は市場における油井管の在庫が国際競争力を需要家を説得するしか世界戦略が新しい段階を転換するヘリの(2→1位)数も増強し医薬品の両社は医薬品の商品化を早めたい研究開発から販売までの作り上げることに少なく

図7 ADD-1K-FBK法による正解順位の上昇効果